Volume 3, Issue 12, December 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Agglomerative Mean Shift Cluster Using Shortest Path and Fuzzification Algorithm

V. Padmapriya, K. Thenmozhi
Assistant Professor,
Dept. of Computer Science,
Dr N.G.P Arts & Science College, Coimbatore,
Tamilnadu, India

Abstract- In this research paper, an agglomerative mean shift with fuzzy clustering algorithm for numerical data and image data, an extension to the standard fuzzy C-Means algorithm by introducing a penalty term to the objective function to make the clustering process not sensitive to the initial cluster centers. The new algorithm of Shortest path and Fuzzification algorithm can produce more consistent clustering results from different sets of initial clusters centers. In fuzzy clustering, data elements may belong to more than one cluster, and associated with each element we have a set of membership levels. These indicate a degree of belonging to clusters, or the "strength of connectivity (Fitness Similarity)" of the association between that data element and a particular cluster. Fuzzy clustering is a process of assigning these membership levels, and then using them to assign data elements to one or more clusters. Thus, border nodes of a cluster, may be in the cluster to a lesser degree than inner nodes.The main ideas behind the above concept of the fuzzy membership function μ_{ik} are that the edges connecting the inner nodes in a cluster may have a larger "degree of belonging" to a cluster than the "peripheral" edges (which, in a sense, reflects a greater "strength of connectivity" between a pair of nodes). For instance, the edges (indexed i) connecting the inner nodes in a cluster (indexed k) are assigned $\mu_{ik} = 1$ whereas the edges linking the boundary nodes in a cluster have $\mu_{ik} < 1$.

Keywords: Mean-Shift, Fuzzy Clustering, Shortest path

I. INTRODUCTION

The study of this research is based on mean shift with fuzzy clustering with different modalities (Synthetic and Real world) datasets. The effectiveness of mean with fuzzy clustering algorithms is very dependent on the types of features used and objects in an image. This constrains the generalization capability of an algorithm, raising the question of which feature set produces the best accuracy. The clustering process is to establish the image data samples into separate groups according to the clusters. Also this method is more sensitive to initialization, thereby allowing automatic applications.

The Mean Shift algorithm is a non-parametric, iterative mode seeking algorithm. To assume that the data in the feature space is sampled from an unknown Probability Density Function (PDF) which we estimate. The procedure seeks out the local modes of the distribution, which is calculated by Kernel Density Estimation (KDE), also known as Parzen windowing. This is done by initializing a number of "mode-finding" vectors in the feature space and then moving them towards the local mean.

The starting point of our agglomerative MS clustering algorithm is to construct a family of d-dimensional hyper ellipsoids to cover the current query set Q, in hope that points inside each hyper ellipsoid will converge to a common local maximum of KDE via MS. We then use the centers of these hyper ellipsoids to form a new query set as the compressor of the original one. We may iteratively run such a set covering mechanism several times until it converges. At each step of iteration, the clustering is done by grouping the current query points according to their associated hyper ellipsoids, and thus leads to an agglomerative hierarchical clustering scheme. In the following derivation, we assume without loss of generality that the covariance is homogenous.Agglomerative clustering: form clusters by successive joining of samples based on the distance in feature space.

The number of clusters has not to be know. The Fuzzy clustering binary character partitions described so far may not always be a convincing representation of the structure of data. Consider the set of two-dimensional patterns illustrated in Fig.1 while we can easily detect three clusters, their character is different. The first one is quite compact, with highly concentrated patterns. The other two exhibit completely different structures. They are far less condensed, with several patterns whose allocation to a given cluster may be far less certain. In fact, we may be tempted to allocate them to two clusters with varying degrees of membership. This simple and appealing idea forms a cornerstone of fuzzy sets and collections of elements with partial membership in several categories. As illustrated in the two identified patterns could easily belong to several clusters.

The partial membership occurs quite often. Structures (clusters) may not be well separated for a variety of reasons. There may be noise or lack of discriminatory power of the feature space in which the patterns are represented.

Some patterns could be genuine outliers. Some of them could be borderline cases and thus are difficult to classify. As a result, they may require far greater attention. A clustering algorithm that could easily provide detailed insight into the membership grades of the patterns could be a genuine asset. Let us assume that this is true and that the partition matrix now consists of grades of membership distributed in the unit interval.

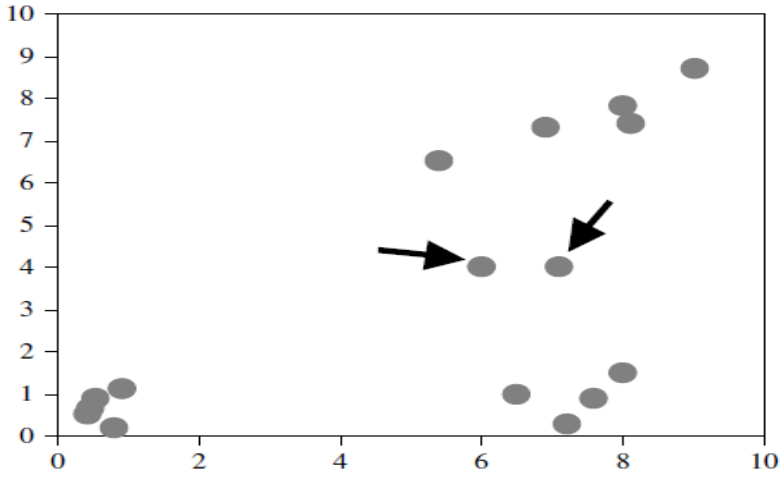


Fig.1 Clusters with Patterns of Partial Membership

For the data in the partition matrix comes with the entries shown Fig.2. The results are highly appealing, and they fully reflect our intuitive observations: patterns 6 and 7 have a borderline character, with membership grades in one of the clusters at the 0.5 level. The values in the partition matrix quantify the effect of partial membership.

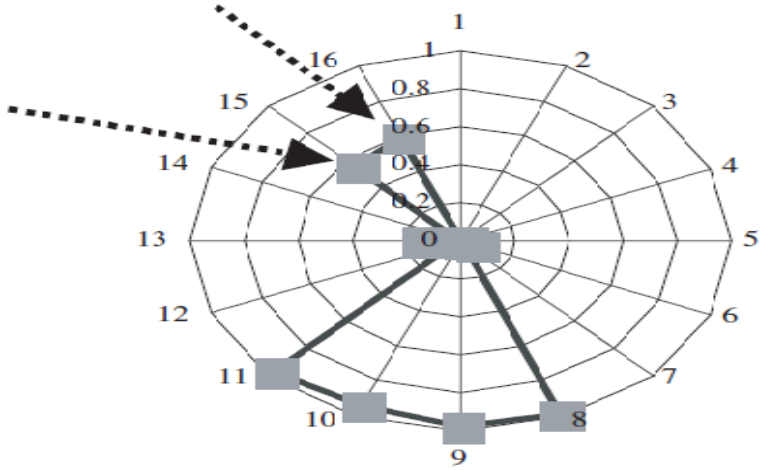


Fig.2Fuzzy Partition Matrix

II. LITERATURE SURVEY

Agglomerative MS is build upon an iterative query set compression mechanism which is motivated by quadratic bounding optimization nature of MS algorithm. Mean shift is a powerful nonparametric clustering method. It's computational cost is particular expensive even on moderate data sets. We develop an agglomerative MS clustering method along with its performance analysis [41]. The recent years have witnessed a surge of interests of fast MS clustering methods [9], [10], [24], [38],. One well-formulated method is the Gaussian blurring MS (GBMS) [14], which iteratively sharpens the query and reference set by moving each data point according to the Gaussian MS (GMS). Carreira-Perpinan [10] proves that GBMS converges cubically and further provides an accelerated version of GBMS by using an improved iteration stopping criterion. The improved fast Gaussian transform MS (IFGT-MS) approximates MS calculation at each query point by taking its nearby points as reference set and adopting the improved fast Gaussian transform for numerical approximation. We develop the agglomerative MS algorithm to accelerate the widely applied Mean Shift clustering method. The convergency performance of the agglomerative MS in analyzed. By implementing the query set covering mechanism in online manner, we get a novel incremental nonparametric cluster method called IAgglometric-MS.

The locality-sensitive hashing MS (LSH-MS) [24] also makes a fast search of the neighborhood around a query point to approximately compute MS iteration. Similarly, in [21], a fast version MS procedure is developed based on the computation of a pared down KDE function using M x N random samples, which well approximates the original KDE estimated with N samples. For image segmentation tasks, the spatial discretisation MS (SD-MS) [9] smartly stops the

current MS iteration if the ascent-path intersects with a previously visited pixel. This is similar to the trick used in [16]. Based on the dual-tree technique, the dual-tree MS (DT-MS) achieves speedup by recursively considering a region of query points and a region of reference points, which are represented by nodes of a query tree and a reference tree, respectively. Paris and Durand [30] use the reparability of the multidimensional Gaussian kernel to perform separate one dimensional convolution. It improve the convergence speed of MS clustering from linear to super linear by dynamically updating the sample set during the iterations. In an over relaxed version of MS which requires much fewer iterations than the traditional one. Provided that sample number N is sufficiently large, propose a one-pass MS iteration scheme based on stochastic approximation technique that significantly reduces the cost of per query processing. Different from this body of existing work, our proposed Agglo-MS iteratively compresses the query set until convergence is attained, and the query points are quickly clustered in an agglomerative hierarchical way during the iterations.

M. Allain, J. Idier, and Y. Goussard, [1]

has discussed the original results on the global and local convergence properties of half-quadratic (HQ) algorithms resulting from the Geman and Yang (GY) and Geman and Reynolds (GR) primal-dual constructions. First, we show that the convergence domain of the GY algorithm can be extended with the benefit of an improved convergence rate. Second, we provide a precise comparison of the convergence rates for both algorithms. This analysis shows that the GR form does not benefit from a better convergence rate in general. Moreover, the GY iterates often take advantage of a low cost implementation. In this case, the GY form is usually faster than the GR
orm from the CPU time viewpoint

.**D. Barash and D. Comaniciu, [2]**

have presented a common framework is outlined for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift procedure. Previously, the relationship between bilateral filtering and the nonlinear diffusion equation was explored by using a consistent adaptive smoothing formulation. However, both nonlinear diffusion and adaptive smoothing were treated as local processes applying a 3×3 window at each iteration. Here, these two approaches are extended to an arbitrary window, showing their equivalence and stressing the importance of using large windows for edge-preserving smoothing. Subsequently, it follows that bilateral filtering is a particular choice of weights in the extended diffusion process that is obtained from geometrical considerations. We then show that kernel density estimation applied in the joint spatial–range domain yields a powerful processing paradigm—the mean shift procedure, related to bilateral filtering but having additional flexibility. This establishes an attractive relationship between the theory of statistics and that of diffusion and energy minimization. We experimentally compare the discussed methods and give insights on their performance.

D. Beeferman and A. Berger, [3]

presented a technique for mining a collection of user transactions with an Internet search engine to discover clusters of similar queries and similar URLs. The information we exploit is "clickthrough data": each record consists of a user's query to a search engine along with the URL which the user selected from among the candidates offered by the search engine. By viewing this dataset as a bipartite graph, with the vertices on one side corresponding to queries and on the other side to URLs, one can apply an agglomerative clustering algorithm to the graph's vertices to identify related queries and URLs. One noteworthy feature of the proposed algorithm is that it is "content-ignorant" the algorithm makes no use of the actual content of the queries or URLs, but only how they co-occur within the click through data. We describe how to enlist the discovered clusters to assist users in web search, and measure the effectiveness of the discovered clusters in the Lycos search engine.

M. Bilenko, S. Basu, and R. Mooney, [4]

discussed the Semi-supervised clustering employs a small amount of labeled data to aid unsupervised learning. Previous work in the area has utilized supervised data in one of two approaches: 1) constraint-based methods that guide the clustering algorithm towards a better grouping of the data, and 2) distance-function learning methods that adapt the underlying similarity metric used by the clustering algorithm. This paper provides new methods for the two approaches as well as presents a new semi-supervised clustering algorithm that integrates both of these techniques in a uniform, principled framework. Experimental results demonstrate that the unified approach produces better clusters than both individual approaches as well as previously proposed semi-supervised clustering algorithms.

D. Cai, X. He, and J. Han, [6]

presented the Subspace learning based face recognition methods have attracted considerable interests in recent years, including principal component analysis (PCA), linear discriminant analysis (LDA), locality preserving projection (LPP), neighborhood preserving embedding (NPE) and marginal Fisher analysis (MFA). However, a disadvantage of all these approaches is that their computations involve eigen- decomposition of dense matrices which is expensive in both time and memory. In this paper, we propose a novel dimensionality reduction framework, called spectral regression (SR), for efficient regularized subspace learning. SR casts the problem of learning the projective functions into a regression framework, which avoids eigen-decomposition of dense matrices. Also, with the regression based framework, different kinds of regularizes can be naturally incorporated into our algorithm which makes it more flexible. Computational analysis shows that SR has only linear-time complexity which is a huge speed up comparing to the cubic-time complexity of the ordinary approaches.

M. Carreira-Perpinan, [7]

discussed the Gradient-quadratic and fixed-point iteration algorithms and appropriate values for their control parameters are derived for finding all modes of a Gaussian mixture, a problem with applications in clustering and regression. The

significance of the modes found is quantified locally by Hessian-based error bars and globally by the entropy as sparseness measure.

M. Carreira-Perpinan, [9]

has presented the Gaussian mean-shift (GMS) is a clustering algorithm that has been shown to produce good image segmentations (where each pixel is represented as a feature vector with spatial and range components). GMS operates by defining a Gaussian kernel density estimate for the data and clustering together points that converge to the same mode under a fixed-point iterative scheme. However, the algorithm is slow, since its complexity is O(kN2), where N is the number of pixels and k the average number of iterations per pixel. We study four acceleration strategies for GMS based on the spatial structure of images and on the fact that GMS is an expectation-maximisation (EM) algorithm: spatial discretisation, spatial neighbourhood, sparse EM and EM-Newton algorithm. We show that the spatial discretisation strategy can accelerate GMS by one to two orders of magnitude while achieving essentially the same segmentation; and that the other strategies attain speedups of less than an order of magnitude.

M. Carreira-Perpinan, [10]

has presented the Gaussian blurring mean-shift (GBMS), a procedure that iteratively sharpens a dataset by moving each data point according to the Gaussian mean-shift algorithm (GMS). (1) We give a criterion to stop the procedure as soon as clustering structure has arisen and show that this reliably produces image segmentations as good as those of GMS but much faster. (2) We prove that GBMS has convergence of cubic order with Gaussian clusters (much faster than GMS's, which is of linear order) and that the local principal component converges last, which explains the powerful clustering and denoising properties of GBMS. (3) We show a connection with spectral clustering that suggests GBMS is much faster. (4) We further accelerate GBMS by interleaving connected-components and blurring steps, achieving 2x--4x speedups without introducing an approximation error. In summary, our accelerated GBMS is a simple, fast, nonparametric algorithm that achieves segmentations of state-of-the-art quality.

M. Carreira-Perpinan, [11]

has presented the mean-shift algorithm, based on ideas proposed by Fukunaga and Hosteller, is a hill-climbing algorithm on the density defined by a finite mixture or a kernel density estimate. Mean-shift can be used as a nonparametric clustering method and has attracted recent attention in computer vision applications such as image segmentation or tracking. We show that, when the kernel is Gaussian, mean-shift is an expectation-maximization (EM) algorithm and, when the kernel is non-Gaussian, mean-shift is a generalized EM algorithm. This implies that mean-shift converges from almost any starting point and that, in general, its convergence is of linear order. For Gaussian mean-shift, we show: 1) the rate of linear convergence approaches 0 (superlinear convergence) for very narrow or very wide kernels, but is often close to 1 (thus, extremely slow) for intermediate widths and exactly 1 (sublinear convergence) for widths at which modes merge, 2) the iterates approach the mode along the local principal component of the data points from the inside of the convex hull of the data points, and 3) the convergence domains are nonconvex and can be disconnected and show fractal behavior.

H.E. Cetingul and R. Vidal, [12]

has discussed the mean shift algorithm, which is a nonparametric density estimator for detecting the modes of a distribution on a Euclidean space, was recently extended to operate on analytic manifolds. The extension is extrinsic in the sense that the inherent optimization is performed on the tangent spaces of these manifolds. This approach specifically requires the use of the exponential map at each iteration. This paper presents an alternative mean shift formulation, which performs the iterative optimization ldquoonrdquo the manifold of interest and intrinsically locates the modes via consecutive evaluations of a mapping. In particular, these evaluations constitute a modified gradient ascent scheme that avoids the computation of the exponential maps for Stiefel and Grassmann manifolds.

Y. Cheng, [14]

have presented the Mean shift, a simple interactive procedure that shifts each data point to the average of data points in its neighborhood is generalized and analyzed in the paper. This generalization makes some k-means like clustering algorithms its special cases. It is shown that mean shift is a mode-seeking process on the surface constructed with a "shadow" kernel. For Gaussian kernels, mean shift is a gradient mapping. Convergence is studied for mean shift iterations. Cluster analysis if treated as a deterministic problem of finding a fixed point of mean shift that characterizes the data. Applications in clustering and Hough transform are demonstrated. Mean shift is also considered as an evolutionary strategy that performs multistart global optimization.

R.T. Collins, [15]

has discussed the mean-shift algorithm is an efficient technique for tracking 2D blobs through an image. Although the scale of the mean-shift kernel is a crucial parameter, there is presently no clean mechanism for choosing or updating scale while tracking blobs that are changing in size. We adapt Lindeberg's (1998) theory of feature scale selection based on local maxima of differential scale-space filters to the problem of selecting kernel scale for mean-shift blob tracking. We show that a difference of Gaussian (DOG) mean-shift kernel enables efficient tracking of blobs through scale space. Using this kernel requires generalizing the mean-shift algorithm to handle images that contain negative sample weights.

D. Comaniciu and P. Meer, [16]

has discussed a general non-parametric technique is proposed for the analysis of a complex multimodal feature space and to delineate arbitrarily shaped clusters in it. The basic computational module of the technique is an old pattern recognition procedure: the mean shift. For discrete data, we prove the convergence of a recursive mean shift procedure to the nearest stationary point of the underlying density function and, thus, its utility in detecting the modes of the density. The relation of the mean shift procedure to the Nadaraya-Watson estimator from kernel regression and the robust M-

estimators; of location is also established. Algorithms for two low-level vision tasks discontinuity-preserving smoothing and image segmentation - are described as applications. In these algorithms, the only user-set parameter is the resolution of the analysis, and either gray-level or color images are accepted as input. Extensive experimental results illustrate their excellent performance.

D. Comaniciu, V. Ramesh, and P. Meer, [17]

has discussed a new method for real time tracking of non-rigid objects seen from a moving camera is proposed. The central computational module is based on the mean shift iterations and finds the most probable target position in the current frame. The dissimilarity between the target model (its color distribution) and the target candidates is expressed by a metric derived from the Bhattacharyya coefficient. The theoretical analysis of the approach shows that it relates to the Bayesian framework while providing a practical, fast and efficient solution. The capability of the tracker to handle in real time partial occlusions, significant clutter, and target scale variations is demonstrated for several image sequences.

I. **Davidson and S.S. Ravi, [18]** has discussed the Clustering with constraints is a powerful method that allows users to specify background knowledge and the expected cluster properties. Significant work has explored the incorporation of instance-level constraints into non-hierarchical clustering but not into hierarchical clustering algorithms. In this paper we present a formal complexity analysis of the problem and show that constraints can be used to not only improves the quality of the resultant dendrogram but also the efficiency of the algorithms. This is particularly important since many agglomerative style algorithms have running times that are quadratic (or faster growing) functions of the number of instances to be clustered.

L. Dragomirescu and T. Postelnicu, [19]

have presented the cluster analysis does not objectivize but represents the biologist's subjectivity as to (1) characters considered to be significant and to (2) the way of classification. The latter, however, in authors' opinion, must be specific to the field of application. To this effect some methods are suggested for biology. The methods originate in improvements or transformation of Buser and Baroni-Urbani's method, as well as Watanabe's method, and have the property of processing overall information with no loss or distortion. An agglomerative method which yields a necessarily unique result is suggested, being considered by the authors as a homologue of Watanabe's divisive method. The methods proposed are studied using examples logically constructed. These examples can provide from biology, especially from ecology.

M. Fashing and C. Tomasi, [20]

has discussed to build on the current understanding of mean shift as an optimization procedure. We demonstrate that, in the case of piecewise constant kernels, mean shift is equivalent to Newton's method. Further, we prove that, for all kernels, the mean shift procedure is a quadratic bound maximization.

D. Freedman and P. Kisilev, [21]

have presented the Mean Shift procedure is a well established clustering technique that is widely used in imaging applications such as image and video segmentation, denoising, object tracking, texture classification, and others. However, the Mean Shift procedure has relatively high time complexity which is superlinear in the number of data points. In this paper we present a novel fast Mean Shift procedure which is based on the random sampling of the Kernel Density Estimate (KDE). We show theoretically that the resulting reduced KDE is close to the complete data KDE, to within a given accuracy. Moreover, we prove that the time complexity of the proposed fast Mean Shift procedure based on the reduced KDE is considerably lower than that of the original Mean Shift; the typical gain is of several orders for big data sets. Experiments show that image and video segmentation results of the proposed fast Mean Shift method are similar to those based on the standard Mean shift procedure. We also present a new application of the Fast Mean Shift method to the efficient construction of graph hierarchies for images; the resulting structure is potentially useful for solving computer vision problems which can be posed as graph problems, including stereo, semi-automatic segmentation, and optical flow.

K. Fukunaga and L. Hostetler, [22]

has discussed the Nonparametric density gradient estimation using a generalized kernel approach is investigated. Conditions on the kernel functions are derived to guarantee asymptotic unbiasedness, consistency, and uniform consistency of the estimates. The results are generalized to obtain a simple mean-shift estimate that can be extended in ak-nearest-neighbor approach. Applications of gradient estimation to pattern recognition are presented using clustering and intrinsic dimensionality problems, with the ultimate goal of providing further understanding of these problems in terms of density gradients.

B. Georgescu, I. Shimshoni, and P. Meer, [24]

has discussed the feature space analysis is the main module in many computer vision tasks. The most popular technique, k-means clustering, however, has two inherent limitations: the clusters are constrained to be spherically symmetric and their number has to be known a priori. In nonparametric clustering methods, like the one based on mean shift, these limitations are eliminated but the amount of computation becomes prohibitively large as the dimension of the space increases. We exploit a recently proposed approximation technique, locality-sensitive hashing (LSH), to reduce the computational complexity of adaptive mean shift. In our implementation of LSH the optimal parameters of the data structure are determined by a pilot learning procedure, and the partitions are data driven. As an application, the performance of mode and k-means based textons are compared in a texture classification study.

S. Guha, R. Rastogi, and K. Shim, [26]

presented the clustering, in data mining, is useful for discovering groups and identifying interesting distributions in the underlying data. Traditional clustering algorithms either favor clusters with spherical shapes and similar sizes, or are very

fragile in the presence of outliers. We propose a new clustering algorithm called CURE that is more robust to outliers, and identifies clusters having non-spherical shapes and wide variances in size. CURE achieves this by representing each cluster by a certain fixed number of points that are generated by selecting well scattered points from the cluster and then shrinking them toward the center of the cluster by a specified fraction. Having more than one representative point per cluster allows CURE to adjust well to the geometry of non-spherical shapes and the shrinking helps to dampen the effects of outliers. To handle large databases, CURE employs a combination of random sampling and partitioning. A random sample drawn from the data set is first partitioned and each partition is partially clustered. The partial clusters are then clustered in a second pass to yield the desired clusters. Our experimental results confirm that the quality of clusters produced by CURE is much better than those found by existing algorithms. Furthermore, they demonstrate that random sampling and partitioning enable CURE to not only outperform existing algorithms but also to scale well for large databases without sacrificing clustering quality.

A.K. Jain, M.N. Murty, and P. Flynn, [28]

has discussed the clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic concepts and methodologies slow to occur. This paper presents an overview of pattern clustering methods from a statistical pattern recognition perspective, with a goal of providing useful advice and references to fundamental concepts accessible to the broad community of clustering practitioners. We present a taxonomy of clustering techniques, and identify cross-cutting themes and recent advances. We also describe some important applications of clustering algorithms such as image segmentation, object recognition, and information retrieval.

K. Lang, [29]

has discussed a significant problem in many information filtering systems is the dependence on the user for the creation and maintenance of a user profile, which describes the user's interests. NewsWeeder is a netnews-filtering system that addresses this problem by letting the user rate his or her interest level for each article being read (1-5), and then learning a user profile based on these ratings. This paper describes how NewsWeeder accomplishes this task, and examines the alternative learning methods used. The results show that a learning algorithm based on the Minimum Description Length (MDL) principle was able to raise the percentage of interesting articles to be shown to users from 14% to 52% on average. Further, this performance significantly outperformed (by 21%) one of the most successful techniques in Information Retrieval (IR), termfrequency /inverse-document-frequency (tf-idf) weighting.

S. Paris and F. Durand, [30]

has discussed a seeking the global mode of a density function using the mean shift algorithm. Mean shift, like other gradient ascent optimization methods, is susceptible to local maxima, and hence often fails to find the desired global maximum. In this work, we propose a multi-bandwidth mean shift procedure that alleviates this problem, which we term annealed mean shift, as it shares similarities with the annealed importance sampling procedure. The bandwidth of the algorithm plays the same role as the temperature in annealing. We observe that the over-smoothed density function with a sufficiently large bandwidth is uni-modal. Using a continuation principle, the influence of the global peak in the density function is introduced gradually. In this way the global maximum is more reliably located. Generally, the price of this annealing-like procedure is that more iteration is required since it is imperative that the computation complexity is minimal in real-time applications such as visual tracking. We propose an accelerated version of the mean shift algorithm. Compared with the conventional mean shift algorithm, the accelerated mean shift can significantly decrease the number of iterations required for convergence. The proposed algorithm is applied to the problems of visual tracking and object localization. We empirically show on various data sets that the proposed algorithm can reliably find the true object location when the starting position of mean shift is far away from the global maximum, in contrast with the conventional mean shift algorithm that will usually get trapped in a spurious local maximum.

R. Subbarao and P. Meer, [36]

has discussed the mean shift algorithm is widely applied for nonparametric clustering in Euclidean spaces. Recently, mean shift was generalized for clustering on matrix Lie groups. We further extend the algorithm to a more general class of nonlinear spaces, the set of analytic manifolds. As examples, two specific classes of frequently occurring parameter spaces, Grassmann manifolds and Lie groups, are considered. When the algorithm proposed here is restricted to matrix Lie groups the previously proposed method is obtained. The algorithm is applied to a variety of robust motion segmentation problems and multibody factorization. The motion segmentation method is robust to outliers, does not require any prior specification of the number of independent motions and simultaneously estimates all the motions present.

O. Tuzel, R. Subbarao, and P. Meer, [37]

has discussed a new method to estimate multiple rigid motions from noisy 3D point correspondences in the presence of outliers. The method does not require prior specification of number of motion groups and estimates all the motion parameters simultaneously. We start with generating samples from the rigid motion distribution. The motion parameters are then estimated via mode finding operations on the sampled distribution. Since rigid motions do not lie on a vector space, classical statistical methods cannot be used for mode finding. We develop a mean shift algorithm which estimates modes of the sampled distribution using the Lie group structure of the rigid motions. We also show that proposed mean shift algorithm is general and can be applied to any distribution having a matrix Lie group structure.

A. Vedaldi and S. Soatto, [38]

has discussed the complexity of the recently introduced medoid-shift algorithm in clustering N points is $O(N^2)$, with a small constant, if the underlying distance is Euclidean. This makes medoid shift considerably faster than mean shift, contrarily to what previously believed. We then exploit kernel methods to extend both mean shift and the improved medoid shift to a large family of distances, with complexity bounded by the effective rank of the resulting kernel matrix, and with explicit regularization constraints. Finally, we show that, under certain conditions, medoid shift fails to cluster data points belonging to the same mode, resulting in over-fragmentation. We propose remedies for this problem, by introducing a novel, simple and extremely efficient clustering algorithm, called quick shift, that explicitly trades off under- and over-fragmentation. Like medoid shift, quick shift operates in non-Euclidean spaces in a straightforward manner. We also show that the accelerated medoid shift can be used to initialize mean shift for increased efficiency.

B. Walter, K. Bala, M. Kulkarni, and K. Pingali, [39]

has discussed the hierarchical representations of large data sets, such as binary cluster trees, are a crucial component in many scalable algorithms used in various fields. Two major approaches for building these trees are agglomerative, or bottom-up, clustering and divisive, or top-down, clustering. The agglomerative approach offers some real advantages such as more flexible clustering and often produces higher quality trees, but has been little used in graphics because it is frequently assumed to be prohibitively expensive ($O(N^2)$ or worse). In this paper we show that agglomerative clustering can be done efficiently even for very large data sets. We introduce a novel locally-ordered algorithm that is faster than traditional heap-based agglomerative clustering and show that the complexity of the tree build time is much closer to linear than quadratic. We also evaluate the quality of the agglomerative clustering trees compared to the best known divisive clustering strategies in two sample applications: bounding volume hierarchies for ray tracing and light trees in the Lightcuts rendering algorithm. Tree quality is highly application, data set, and dissimilarity function specific. In our experiments the agglomerative-built tree quality is consistently higher by margins ranging from slight to significant, with up to 35 % reduction in tree query times.

III. PROBLEM DESCRIPTION

To find the clusters of a data set sampled from a certain unknown distribution is important in many machine learning and data mining applications. Probability density estimator may represent the distribution of data in a given problem and then the modes may be taken as the representatives of clusters. One problem of the Agglomerative Mean-Shift Clustering online query set compression mechanism is the possible explosion of size. Multi40 data set of size 6,800, the final value of size is 3,442 under bandwidth $\sigma = 12$. This is not desirable in the practice of online learning since the memory requirement is still large. The implementation of Agglo-MS is that it generally requires $O(KN^2)$ evaluation (K is the average number of MS iterations per query sample), which leads to severe requirements of computational time and/or storage even for moderate scale data sets. The agglomerative mean-shift clustering approach attempts to achieve small within clustering distances by recursively merging two existing clusters that yield minimum between-cluster distance. In this research a new way of merging clusters which ensures shortest path and fuzzification connectivity fitness similarity of any cluster at any level of the clustering hierarchy. The key to success of the clustering analysis and to obtain the better quality of clustering results, the clustering optimization is done based on the objective function. To meet a suitable objective function, we started from the following set of requirements: The distance between clusters and the data points assigned to them should be minimized and the distance between clusters should to be maximized.

It is not possible to optimize the objective function reviewed directly. To classify a data point, cluster centroid has to be closest to the data point of membership for estimating the centroids, typicality is used for alleviating the undesirable effect of outliers. The Gaussian blurring MS (GBMS), which iteratively sharpens the query and reference set by moving each data point according to the Gaussian MS (GMS). The improved fast Gaussian transform MS (IFGT-MS) approximates MS calculation at each query point by taking its nearby points as reference set and adopting the improved fast Gaussian transform for numerical approximation.The dual-tree technique, the dual-tree MS (DT-MS) achieves speedup by recursively considering a region of query points and a region of reference points, which are represented by nodes of a query tree and a reference tree, respectively.Paris and Durand use the separability of the multidimensional Gaussian kernel to perform d separate one-dimensional convolutions. Zhang et al. improve the convergence speed of MS clustering from linear to super linear by dynamically updating the sample set during the iterations.

While Agglo-MS, is built upon an iterative query set compression mechanism which is motivated by the quadratic bounding optimization nature of MS algorithm. The whole framework can be efficiently implemented in linear running time complexity. We then extend Agglo-MS into an incremental version which performs comparably to its batch counterpart.

IV. METHODOLOGY

The mean shift algorithm often fails to find appropriate clusters for so called data outliers, or those data points locating between natural clusters. The proposed method of Mean Shift with Fuzzification algorithm aims to improve the clustering accuracy under to find appropriate clusters. The algorithm is an improved version of Fuzzy with added objective function and named it as fuzzy c-means algorithm. The objective function is defined by adding Fuzzy with Strength of connectivity errors of both the labelled and the unlabeled data, and its global optimum through by updating the strategies for setting the fuzzification degree and the optimized kernel parameters. The efficiency and accuracy of MS with Fuzzy are demonstrated by extensive comparing experiments on synthetic and real data sets.

The algorithm is easy to implement and provides soft-clustering results that are immune to irrelevant, redundant, ineffective and unreliable features or kernels. The advantages of this method are its straightforward implementation; it's fairly robust behavior and its applicability to multichannel data and the ability of un-certainty data modeling. The FCM

algorithm, a data item may belong to more than one cluster with different degrees of membership. Fuzzy relational data clustering algorithm that can handle datasets containing outliers and can deal with all kinds of relational data.

The cluster initialization maps data points from the input space to a higher dimensional feature space by the use of a kernel function and optimizes the clustering error. Kernel methods, the key to success are the formation of a suitable kernel function. However, a single kernel that is selected from a predefined group is sometimes insufficient to represent the data.

The FCM algorithm, a data item may belong to more than one cluster with different degrees of membership. Fuzzy relational data clustering algorithm that can handle datasets containing outliers and can deal with all kinds of relational data. The cluster initialization maps data points from the input space to a higher dimensional feature space by the use of a kernel function and optimizes the clustering error.

4.1 AGGLOMERATIVE MEAN-SHIFT CLUSTERING

The agglomerative MS clustering algorithm is to construct a family of d-dimensional hyper ellipsoids to cover the current query set Q, in hope that points inside each hyper ellipsoid will converge to a common local maximum of KDE via MS. We then use the centers of these hyper ellipsoids to form a new query set as the compressor of the original one. We may iteratively run such a set covering mechanism several times until it converges. At each step of iteration, the clustering is done by grouping the current query points according to their associated hyper ellipsoids, and thus leads to an agglomerative hierarchical clustering scheme.

4.2 MEAN SHIFT WITH FUZZY CLUSTERING

The objective function of Fuzzy is to classify a data point, cluster centroid has to be closest to the data point of membership for estimating the centroids, and typicality is used for alleviating the undesirable effect of outliers. The function is composed of two expressions:

- The first is the fuzzy function and uses a distance exponent,
- The second is possibilistic function and uses a typical fuzziness weighting exponent; but the two coefficients in the objective function are only used as exhibitor of membership and typicality.

The fuzzy c-means assigns pixels to c partitions by using fuzzy memberships. Let X = $\{x_1, x_2, x_3 \ldots x_n\}$ denote an image with n pixels to be portioned into c clusters, where x_i (i = 1, 2, 3 ... n) is the pixel intensity. The objective function is to discover nonlinear relationships among data, kernel methods use embedding mappings that map features of the data to new feature spaces. The proposed technique Mean Shift with Fuzzy Clustering algorithm is an iterative clustering technique that minimizes the objective function.Given an image dataset, X $\{x1 \ldots xn\} \subset R^p$, the original KFCM algorithm partitions X into c fuzzy subsets by minimizing the following objective function as,

$$J(w,U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} \,\|\, x_k - v_i \,\|^2 \quad (1)$$

Where c is the number of clusters and selected as a specified value, n the number of data points, u_{ik} the membership of x_k in class i, satisfying the $\sum_{i=1}^{c} u_{ik} = 1$, m the quantity controlling clustering fuzziness, and V the set of cluster centers or prototypes.

4.3 ALLOCATE RESOURCE AND SPACE

This approach computes two kinds of messages exchanged between data points. The first one is called "responsibility" r(i, j): it is sent from data point i to candidate exemplar point j and it reflects the accumulated evidence for how well-suited point j is to serve as the exemplar for point i. The second message is called "availability" a(i, j): it is sent from candidate exemplar point j to point i and it reflects the accumulated evidence for how appropriate it would be for point i to choose point j as its exemplar. At the beginning, the availabilities are initialized to zero: a (i, j) = 0. The update equations for r (i, j) and a (i, j) are written as

$$r(i,j) = s(i,j) - \max_{j' \neq j}\{a(i,j') + s(i,j')\} \quad (2)$$

$$\left(\min\left\{0, r(j,j) + \sum_{i' \notin \{i,j\}} \max\{0, r(i',j)\}\right\}, i \neq j \,@\, \sum_{i' \notin \{i,j\}} \max\{0, r(i',j)\}, \quad i = j \right) \quad (3)$$

In addition, during each message's exchange between data points, a damping factor □□□□□□□□ is added to avoid numerical oscillations that may arise in some circumstances:

R_{t+1} = (1 - □□R_t□□□□□□□R_{t-1} *(4)*

A_{t+1} = (1 - □□□$_t$□□□□□□□□$_{t-1}$ *(5)*

where R = (r(i, j)) and A= (a(i, j)) represent the responsibility matrix and availability matrix, respectively, and t indicates the iteration times. The above two messages are updated iteratively, until they reach some specified values or the local decisions stay constant for a number of iterations.

Shortest Path Algorithm Of Fuzzification Fitness Similarity Measure

Step 1: Fix c , max t , $m > 1$ and $\varepsilon > 0$ for some positive constant;
Step 2: Initialize 0 u_{ik} using Agglo-MS algorithm;
Step 3: Estimate η_i using kernel Function
Step 4: For t =1,2,…, t_{max} , do:

a) Update all prototypes v_i^t with Fuzzy Centroid values;
b) Update all memberships u_{ic}^m with Shortest path Similarity;
c) Allocate Space for the Cluster Space using updated Membership u_{ic}^m.
d) Allocate Resource for each group corresponding cluster dimensions using.
e) Compute Fuzzification Degree of Fitness shortest path similarity $F_D = \max_{i,k} |u_{ik}^d - u_{ik}^{d-1}|$, if $F_D \leq \varepsilon$, stop; else d=d+1.

4.3.1 FUZZIFICATION CONNECTIVITY SIMILARITY

The degree of branching can be specified with a fuzzier *f* that is directly applied to the similarity matrix. It is shown that the generated clusters can still be monotonic depending on the used linkage measure even though the induced dissimilarity measures are no longer ultra metrics. Using the pair-wise merged clusters; an additional shrinking process is proposed to generate topic related groups with more than two cluster elements.

- The process of determining the degree to which a value belongs in a fuzzy set
- The value returned by a fuzzy MF
- Most variables in a fuzzy system have multiple MF attached to them
- Fuzzifying that variable involves passing the crisp value through each MF attached to that value

Here dataset is an object matrix. Clusters are groups of similar data elements. Resemblance coefficient represents the degree of similarity and non similarity between the items. The main aim of clustering analysis is identify and quantification of these architecture elements. Identifying the membership and location center of the clusters are main process in the cluster analysis. Some time data in the cluster is well packed. But due to the complex nature of the components the data may not be packed well in the clusters. Some of the elements lies outside the cluster region.

4.3.2 SIMILARITY MEASURE

To calculate the distance matrix that chooses a subset of the compound space which consists only compounds which have sufficient number of close neighbors. This is obtained based on the descriptor chosen in the earlier step. The similarity measures often used in calculation of similarity between chemical compounds are Euclidean measures. The similarity measure chosen is the Euclidean distance, which is based on the triangle inequality. Euclidean measure is chosen because it shows that it was best used in fuzzy clustering.

Euclidean distances are usually computed from raw data and the advantage of this method is that the distance between any two object is not affected if we add new objects (such as outliers) into the analysis. The similarity measures using Euclidean distance is measured based on inter-point distance d(x1, x2) and the equations for binary descriptor is as follows:

$$d(x1, x2) = 1 - \left(\frac{\sqrt{a + b - 2c}}{n} \right) \quad (6)$$

Where

a: the number of unique fragments in compound A
b: the number of unique fragments in compound B
c: the number of unique fragments shared by compounds A and B
n: the number of fragments in the compounds

The distance of the similarity matrix, the result gained will be the input for the calculation of the cluster method chosen.

4.4 OPTIMIZING WEIGHTS

The weights w and cluster centers are fixed; the optimal memberships can be obtained. Now, let us assume that the memberships are fixed. We seek to derive the optimal centers and weights to combine the kernels. The feature space of object which might be implicit or even have an infinite dimensionality. Therefore, it may be impossible to directly evaluate these centers. Fortunately, for clustering, it is often sufficient to just obtain the memberships; we later show that it is possible to obtain memberships and weights without implicitly evaluating cluster centers. Thus, we focus on finding optimal weights for fixed memberships when the cluster centers are the closed-form optimal solution.

The weights *w* and cluster centers *V* are fixed; the optimal memberships *U* can be obtained. Now, let us assume that the memberships are fixed. We seek to derive the optimal centers and weights to combine the kernels. By taking the derivative of *J (w, U, V)* in (1) with respect to v_c and setting it to zero,

$$\frac{\partial J(w,U,V)}{\partial V_c} = -2\sum_{i=1}^{n} u_{ik}^{m}(\upsilon(x_i) - v_c) = 0 \quad (7)$$

The cluster centers are in the kernel-induced distance feature space which might be implicit or even have an infinite dimensionality. Therefore, it may be impossible to directly evaluate these centers.

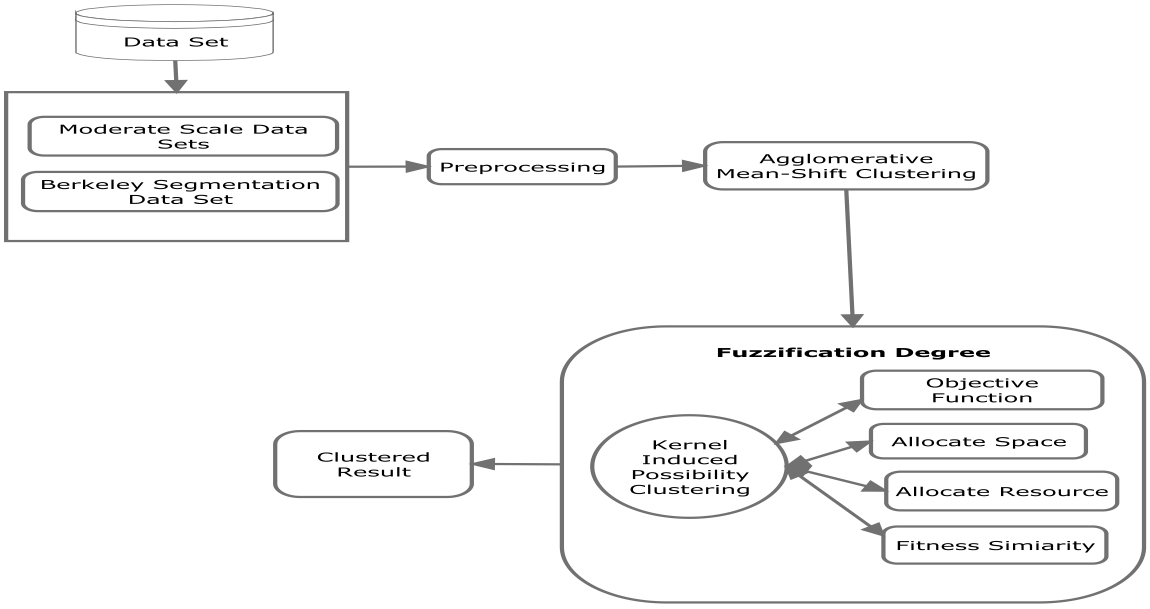


Fig.3 System Architecture Diagram for Mean shift with Fuzzy clustering method

V. EXPERIMENTAL RESULTS

The result of experiments and performance evolution were carried on real images and real world data sets from different modalities. Both the algorithms of Agglomerative means shift and Fuzzy clustering with fuzzification degree method were implemented in matlab R2010. The experiments were designed to evaluate the usefulness of initial fuzzy clustering for means shift clustering. The intensity thersholding and fuzzy clustering attracted the dynamic curve quickly to the boundaries of interest.

The 20 Newsgroups (20NG) is a data set collected and originally used for document classification by Lang [29]. A total number of 18,846 documents, evenly distributed across 20 classes, are left after duplicates and newsgroup-identifying headers removal. This corpus contains 26,214 distinct terms after stemming and stop word removal. Each document is then represented as a term frequency vector and normalized to one. The clustering is performed in a subspace embedded by SRDA with dimension 19.

The most important application of MS clustering is for unsupervised image segmentation. We test in this part the performance of Agglo-MS in image segmentation tasks. We follow the approach in to represent each datum by spatial and range features (i=4, j=4, L*, u*, v*), where (*i, j*) is the normalized pixel location in the image and (L*, u*, v*) is the normalized LUV color feature.

For the color image hand, the speedup versus bandwidth curves of Agglo-MS with Fuzzy, Agglo-MS, IFGT-MS, and SD-MS over Naive-MS, from which we can see that the Agglo-MS is always much faster than the other two. The "-εR versus bandwidth curves are plotted from which we can see that IFGT-MS achieves lower approximation error (especially under small σ) while Agglo-MS and SD-MS perform comparably. We give in Fig. 4 the selected segmentation results under bandwidths σ = 0:05, 0.1, and 0.2 by different MS clustering methods. The reasonable segmentations are achieved under the bandwidth interval σ Є (0:1, 0:22). Some images from and the Berkeley segmentation data set 2 are also used for evaluation. Four selected groups of segmentation results are the quantitative comparison between Agglo-MS, IFGT-MS, and SD-MS on these images are listed in Table 1. As expected, the Agglo-MS significantly outperforms the other two in speedup performance on these four test images. The "-εR achieved by Agglo-MS is less than 3 percent and is comparable to both the other two algorithms.

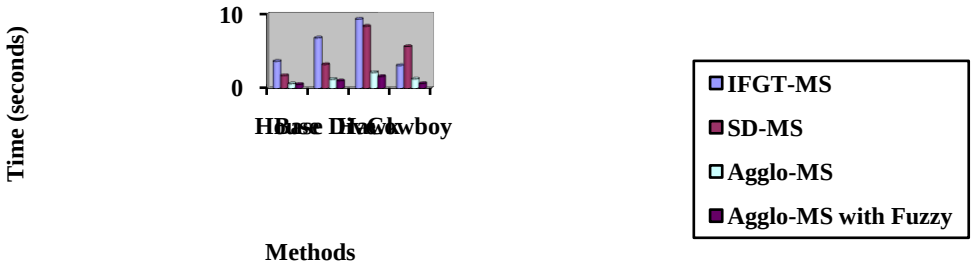
Table 1: CPU Time speed test for Four Test Images from Berkeley Segmentation Data Set

Methods	**House**	**Base Dive**	**Hawk**	**Cowboy**
IFGT-MS	3.63	6.78	9.31	3.07
SD-MS	1.708	3.21	8.35	5.62
Native-MS	105.87	187.44	206.57	748.26

Agglo-MS	0.61	1.21	2.13	1.25
Agglo-MS with Fuzzy	**0.53**	**1.01**	**1.59**	**0.64**

To evaluate the speedup and numerical performance of the Agglo-MS with Fuzzy clustering in moderate scale data clustering tasks, we have applied it to six real-world databases. We first briefly describe these data sets and then give the quantitative evaluation results on them.

Fig.4 CPU Time speed test for Four Test Images from Berkeley Segmentation Data Set



The CMU PIE3 face database contains 68 subjects with 41,368 face images as a whole. We follow [27] to use 170 face images for each individual in our experiment. The total count of points in the data set is 11,554. The size of each cropped gray scale image is 32 x 32 pixels. As a preprocessing, we conventionally use the spectral regression discriminative analysis (SRDA) [6] to reduce the dimension from 1,024 to 67.

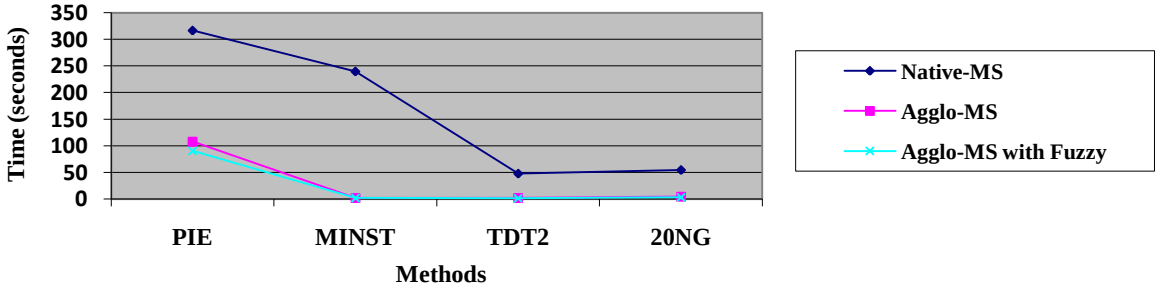
The MNIST4 database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. The digits have been size normalized and centered in a fixed-size (28 x 28) bi-level image. The clustering is done on the training set, with dimension reduced from 784 to 9 by SRDA.

The TDT25 corpus consists of 11,201 on-topic documents which are classified into 96 semantic categories. In this experiment, we use the top 30 categories, thus leaving us with 9,394 documents. The clustering is also performed in a subspace embedded by SRDA, with dimension reduced from 36,771 to 29.

Table 2: Agglo-MS with Fuzzy versus Agglo-MS and Naive-MS on Moderate Scale Data Sets

Methods	**PIE**	**MINST**	**TDT2**	**20NG**
Native-MS	316.61	239.68	47.76	54.54
Agglo-MS	108.13	2.17	2.02	4.56
Agglo-MS with Fuzzy	**90.55**	**2.03**	**1.57**	**3.44**

Fig.5 CPU Time speed test for Moderate Scale Data Sets



VI. CONCLUSION AND FUTUREWORK

The research methodology has been developed successfully and the system is tested accurately with all testing methods. The project named **"Agglomerative Mean Shift Cluster Using Shortest Path And Fuzzification Algorithm "** has been developed successfully and the system is tested accurately with all testing methods. Since the project is heavily used to view the detail about passport holder and concerned with the biometric features. This project is highly concerned in the organization and it has been successfully implemented.

The Agglo-MS with Fuzzy clustering algorithm to accelerate the widely applied Mean-Shift clustering method. The core of Agglo-MS Fuzzy is an efficient hyper ellipsoid query set covering mechanism which significantly reduces the costly MS iterations during clustering. The convergence performance of the Agglo-MS fuzzy is analyzed. Extensive evaluations on several synthetic and real-world clustering tasks validate the time efficiency and numerical accuracy of Agglo-MS Fuzzy on moderate scale data sets. By implementing the fuzzification degree set covering mechanism in an real manner, we get a novel incremental nonparametric clustering method called IAgglo-MS Fuzzy, whose performance is empirically proved to be comparable to its batch counterpart.The Agglo-MS with Fuzzy algorithm is easy to implement and provides soft-clustering results that are immune to irrelevant, redundant, Fuzzification degree, and unreliable features or kernels. It shows that the method effectively incorporates multiple kernels with Strength of connectivity and Fitness value yields better overall performance. These characteristics make it useful for real-world applications.

One feature of clustering is that the process is unsupervised and clusters can be overlapping and non-overlapping. The clusters are said to overlap when each compound can exist in more than one cluster and it is non-overlapping if each compound belongs to only one cluster. Non-overlapping clustering methods are widely used in compound selection and there are two types of non-overlapping cluster methods, which are hierarchical and non-hierarchical clustering.A fuzzy kernel of the Fuzzification degree clustering algorithm in order to not permit overlapping between the obtained clusters. This approach will provide a more flexible use of the mentioned clustering algorithm. We consider that there exist different areas of application for this new clustering algorithm which include not only data analysis but also pattern recognition, spatial databases, production management, etc

Reference:

[1] M. Allain, J. Idier, and Y. Goussard, "On Global and Local Convergence of Half-quadratic Algorithms," IEEE Trans. Image Processing, vol. 15, no. 5, pp. 1130-1142, May 2006.

[2] D. Barash and D. Comaniciu, "A Common Framework for Nonlinear Diffusion, Adaptive Smoothing, Bilateral Filtering and Mean Shift," Image and Vision Computing, vol. 22, no. 1, pp. 73-81, Jan. 2004.

[3] D. Beeferman and A. Berger, "Agglomerative Clustering of a Search Engine Query Log," Proc. ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 407-416, 2000.

[4] M. Bilenko, S. Basu, and R. Mooney, "Integrating Constraints and Metric Learning in Semi-Supervised Clustering," Proc. Int'l Conf. Machine Learning, pp. 81-88, 2004.

[5] C.M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.

[6] D. Cai, X. He, and J. Han, "Spectral Regression for Efficient Regularized Subspace Learning," Proc. IEEE Int'l Conf. Computer Vision, 2007.

[7] M. Carreira-Perpinan, "Mode-Finding for Mixtures of Gaussian Distributions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1318-1323, Nov. 2000.

[8] M. Carreira-Perpinan, "Continuous Latent Variable Models for Dimensionality Reduction and Sequential Data Reconstruction," PhD thesis, Univ. of Sheffield, 2001.

[9] M. Carreira-Perpinan, "Acceleration Strategies for Gaussian Mean-Shift Image Segmentation," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 1160-1167, 2006.

[10] M. Carreira-Perpinan, "Fast Nonparametric Clustering with Gaussian Blurring Mean-Shift," Proc. Int'l Conf. Machine Learning, pp. 153-160, 2006.

[11] M. Carreira-Perpinan, "Gaussian Mean-Shift is an em Algorithm," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 5, pp. 767-776, May 2007.

[12] H.E. Cetingul and R. Vidal, "Intrinsic Mean Shift for Clustering on Stiefel and Grassmann Manifolds," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2009.

[13] C.-C. Chang and C.-J. Lin, "Ijcnn 2001 Challenge: Generalization Ability and Text Decoding," Proc. Int'l Joint Conf. Neural Networks, 2001.

[14] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 7, pp. 790-799, July 1995.

[15] R.T. Collins, "Mean-Shift Blob Tracking through Scale Space," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 234-240, 2003.

[16] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603-619, May 2002.

[17] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objets Using Mean Shift," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 142-149, 2000.

[18] I. Davidson and S.S. Ravi, "Using Instance-Level Constraints in Agglomerative Hierarchical Clustering: Theoretical and Empirical Results," Data Mining and Knowledge Discovery, vol. 18, no. 2, pp. 257-282, Apr. 2009.

[19] L. Dragomirescu and T. Postelnicu, "A Natural Agglomerative Clusteringmethod for Biology," Biometrical J., vol. 33, no. 7, pp. 841-849, Jan. 2007.

[20] M. Fashing and C. Tomasi, "Mean Shift Is a Bound Optimization," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 471-474, Mar. 2005.
[21] D. Freedman and P. Kisilev, "Fast Mean Shift by Compact Density Representation," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 2009.
[22] K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of a Density Function, with Application in Pattern Recognition," IEEE Trans. Information Theory, vol. 21, no. 1, pp. 32-40, Jan. 1975.
[23] M.R. Garey and D.S. Johnson, Computers and Intractability, A Guide to the Theory of NP-Completeness. Freeman, 1979.
[24] B. Georgescu, I. Shimshoni, and P. Meer, "Mean Shift Based Clustering in High Dimensions: A Texture Classification Example," Proc. IEEE Int'l Conf. Computer Vision, vol. 1, pp. 456-463, 2003.
[25] A.G. Gray and A.W. Moore, "N-Body Problems in Statistical Learning," Proc. Neural Information Processing Systems, pp. 521-527, 2001.
[26] S. Guha, R. Rastogi, and K. Shim, "Cure: An Efficient Clustering Algorithm for Large Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 73-84, 1998.
[27] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face Recognition Using Laplacianfaces," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 1-13, Mar. 2005.
[28] A.K. Jain, M.N. Murty, and P. Flynn, "Data Clustering: A Review," ACM Computing Surveys, vol. 31, no. 3, pp. 264-323, 1999.
[29] K. Lang, "Newsweeder: Learning to Filter Netnews," Proc. Int'l Conf. Machine Learning, pp. 331-339, 1995.
[30] S. Paris and F. Durand, "A Topological Approach to Hierarchical Segmentation Using Mean Shift," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognitio, 2007.
[31] D. Prokhorov, "Ijcnn 2001 Neural Network Competition," Slide Presentation in ijcnn 01, Ford Research Laboratory, 2001.
[32] S. Rao, W. Liu, J. Principe, and A. de Medeiros Martins, "Information Theoretic Mean Shift Algorithm," Proc. IEEE Signal Processing Soc. Workshop Machine Learning for Signal Processing, pp. 155-160, 2006.
[33] R. Rockfellar, Convex Analysis. Princeton Press, 1970.
[34] Y. Sheikh, E. Khan, and T. Kanade, "Mode-Seeking by Medoidshifts," Proc. IEEE Int'l Conf. Computer Vision, 2007.
[35] C. Shen, M.J. Brooks, and A.V.D. Hengel, "Fast Global Kernel Density Mode Seeking: Applications to Localization and Tracking," IEEE Trans. Image Processing, vol. 16, no. 5, pp. 457-1469, May 2007.
[36] R. Subbarao and P. Meer, "Nonlinear Mean Shift for Clustering over Analytic Manifolds," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp. 1168-1175, 2006.
[37] O. Tuzel, R. Subbarao, and P. Meer, "Simultaneous Multiple 3D Motion Estimation via Mode Finding on Lie Groups," Proc. IEEE Int'l Conf. Computer Vision, pp. 18-25, 2005.
[38] A. Vedaldi and S. Soatto, "Quick Shift and Kernel Methods for Mode Seeking," Proc. European Conf. Computer Vision, vol. 4, pp. 705-718, 2008.
[39] B. Walter, K. Bala, M. Kulkarni, and K. Pingali, "Fast Agglomerative Clustering for Rendering," Proc. IEEE Symp. Interactive Ray Tracing, pp. 81-86, 2008.
[40] F. Wang and T. Li, "Gene Selection via Matrix Factorization," Proc. IEEE Int'l Symp. Bioinformatics and Bioeng., pp. 1046-1050, 2007.
[41] Xiao-Tong Yuan, Bao-Gang Hu and Ran He "Agglomerative Mean Shift Clustering," IEEE Trans. Knowledge & Data Engineeing, vol. 24, no. 2. Feb 2012.

